ARI Research Note 98-26

The Long-Term Retention of Knowledge and Skills

Alice F. Healy, Deborah M. Clawson, Danielle S. McNamara, William R. Marmie, Vivian I. Schneider, Timothy C. Rickard, Robert J. Crutcher, Cheri L. King, K. Anders Ericsson, and Lyle E. Bourne Jr.
University of Colorado

Research and Advanced Concepts Office
Michael Drillings, Chief

July 1998

U.S. Army Research Institute
for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Research accomplished under contract for the Department of the Army

University of Colorado

Technical review by

Tony Papa

NOTICES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1998	Annual 5/1/92 - 4/30/93

4. TITLE AND SUBTITLE
The Long-Term Retention of Knowledge and Skills

5. FUNDING NUMBERS
PE 0601102A
2O161102B74F
TA 2901
WU C07
Contract No. MDA903-90-K-0066

6. AUTHOR(S)
Alice Healy, Deborah Clawson, Danielle McNamara, William Marmie, Vivian Schneider, Timothy Rickard, Robert Crutcher, Cheri King, K. Anders Ericcson, & Lyle Bourne, Jr. (University of Colorado)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Colorado
Department of Psychology
Campus Box 345
Boulder CO 80309-0345
(303)492-5032

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Institute
ATTN: PERI-BR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Research Note 98-26

11. SUPPLEMENTARY NOTES
COR: Dr. George Lawton

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution is Unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*): We received three classes of guidelines we found to optimize long-term retention. The first class concerned ways to optimize the conditions of training. We discussed three general guidelines in this class. The first concerned the contextual interference found, for example, with random sequences of tasks as opposed to fixed or predictable sequences. The second concerned training parts of a task versus the whole task. The third concerned the distinction between generating and reading. The second class of guidelines concerned ways to optimize the strategies used. We found that in tasks that require deliberate retrieval from memory, training that promotes efficient encoding strategies maximizes long-term retention. The third class of guidelines concerned ways to attain direct access, or automatic retrieval, from memory. We found in several domains that achieving automaticity requires extensive practice. Further, even when retrieval appears automatic after extensive practice, mediators may still continue to exert their influence.

14. SUBJECT TERMS
long-term retention skill specificity skill maintenance part-whole training

15. NUMBER OF PAGES
58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18 298-102

90002604?3963

The Long-Term Retention of Knowledge and Skills

Alice F. Healy, Deborah M. Clawson, Danielle S. McNamara,
William R. Marmie, Vivian I. Schneider, Timothy C. Rickard,
Robert J. Crutcher, Cheri L. King, K. Anders Ericsson,
and Lyle E. Bourne, Jr.

Department of Psychology, University of Colorado, Boulder

Send correspondence to:

Dr. Alice F. Healy

Department of Psychology

University of Colorado

Campus Box 345

Muenzinger Psychology Building

Boulder, CO 80309-0345

Telephone: 303-492-5032

Table of Contents

I. Features of our Research Program

II. Specificity of Training

III. Guidelines for Improving Long-Term Retention

A. Contextual Interference in Acquisition of Logic Rules

B. Part-Whole Training in Morse Code Reception

C. Part-Whole Training of Tank Gunner Skills

D. A Generation Advantage for Mental Arithmetic and Vocabulary Learning

E. Direct and Mediated Retrieval in Mental Arithmetic

F. Direct and Mediated Retrieval in Vocabulary Acquisition

G. Automatic Processing in Color-Word Interference

IV. Summary and Conclusions

The Long-Term Retention of Knowledge and Skills

Alice F. Healy, Deborah M. Clawson, Danielle S. McNamara, William R. Marmie, Vivian I. Schneider, Timothy C. Rickard, Robert J. Crutcher, Cheri L. King, K. Anders Ericsson, and Lyle E. Bourne, Jr.

For the last seven years we have been engaged in a research program aimed generally at understanding and improving the long-term retention of knowledge and skills. Our initial work (see Healy, Fendrich, Crutcher, Wittman, Gesi, Ericsson, & Bourne, 1992, for a summary) led us to propose that a crucial determinant of retention performance concerns the extent to which procedures acquired during study can be reinstated at test. That is, to demonstrate durable retention across a long delay interval, it is critical that the procedures used when acquiring the knowledge or skill are reinstated at a later time. Using this work as a foundation, we have tried to develop more general guidelines concerning training methods optimal for promoting superior long-term retention. As discussed below, the approach we have taken differs from that used in most earlier studies (see, e.g., Farr, 1987, for a cogent review).

I. Features of our Research Program

Five features of our program together distinguish it from earlier research on retention of knowledge and skills. First, we have been explicitly concerned with optimizing performance after a delay interval rather than inferring superior retention from optimized performance during acquisition (see Schmidt & Bjork, 1992, for a recent discussion of this issue). Toward this end, we are striving to find conditions of training that will enable performance to stand up over time, recognizing that efficiency of training is also a consideration (i.e., optimal training may be costly in terms of the time required). As real-

life experience suggests, optimizing performance after a delay is crucial. In fields such as emergency care and the military (see, e.g., Wisher, Sabol, Sukenik, & Kern, 1991), people often have to assume their duties at short notice and with inadequate opportunities to refresh their skills before they are needed in a life-or-death situation. In this respect we have been guided by Bahrick's (1984) concept of "permastore," a kind of memory that shows great durability over extended time periods as long as several decades. Our goal has been to identify conditions of learning or characteristics of learned material that differentiate between items that do or do not achieve permanency in memory.

Second, relative to most other empirical programs, we use longer retention intervals, usually including tests after several weeks or months, and in some cases including intervals up to one or two years.

Third, we employ a combination of structural and analytic experimental procedures. The structural approach aims to identify and describe the components of specific skills. Toward this end, existing experimental methods are refined and adapted to assess the retention characteristics of skill components after long periods of disuse. The analytic approach is concerned with the experimental investigation of factors influencing and promoting retention. This methodology is used to check hypotheses concerning the characteristics that distinguish between permanent and nonpermanent components of knowledge and skill.

Fourth, we have chosen to conduct comparable experiments over a wide range of different skills and paradigms, under the assumption that theoretical conclusions may rely heavily on the specific nature of the tasks under consideration and in order to capitalize on different processes crucial to retention that can be highlighted in different tasks. Our goal is to identify training guidelines that are either common (general over tasks) or idiosyncratic (specific to a particular task) but stable.

Fifth, we have used a nontraditional method to assess retention. In the traditional study, investigators require all subjects to achieve a fixed criterion of performance mastery

in terms of accuracy. Retention is assessed by examining changes in the percentage of subjects who maintain that accuracy criterion as a function of delay. Farr (1987) criticized this traditional approach, suggesting that there are many other factors that can influence retention beyond reaching some mastery criterion (see Underwood, 1964, for another cogent discussion of this issue). For a variety of reasons, we have developed a method which differs in several important respects from the traditional approach. First, we provide training for subjects beyond the accuracy criterion. Second, especially when accuracy measures are near the ceiling, we monitor aspects of the skill that reveal performance changes beyond those evident by assessing accuracy alone, for example, component response time (RT) measures and verbal protocols. These measures provide us a means for defining overlearning without resorting simply to the number of trials after the accuracy criterion has been reached. These additional measures are used to assess retention performance as well as acquisition performance.

This research has led to the support or identification of several guidelines for improving long-term retention of skills. Initially we will state these ideas in general terms; then we will provide evidence for them in terms that are specific to particular experimental paradigms. Each of these guidelines should have application to a host of tasks, as is illustrated by the many different tasks studied in our research program. In this chapter, we focus on three classes of guidelines: those that relate to (a) optimizing conditions of training, (b) optimizing the learning strategy used, and (c) training to achieve automatic levels of processing.

In our earlier studies, we were impressed with the remarkable degree of long-term retention that subjects were able to achieve in a number of perceptual, cognitive, and motor tasks, including studies of target detection, data entry, and mental arithmetic (see Fendrich, Healy, & Bourne, 1991; Fendrich, Healy, & Bourne, in press; Healy et al., 1992; Healy, Fendrich, & Proctor, 1990). Our more recent research has helped to clarify the limits of

this durable retention phenomenon, and we will present some evidence on those limitations before we discuss the optimization guidelines.

II. Specificity of Training

Two of the most significant questions one can ask about the effects of any training program are (a) how general and (b) how durable are these effects? Optimal training programs are those for which effects can be shown to be both general over a range of new situations within a given task domain and durable in the sense that performance suffers minimally over periods of disuse. In fact, however, training effects are often limited to the situations encountered during training and subject to significant forgetting in time (see Gick & Holyoak, 1987, for a discussion of this issue). Our evidence bearing on the reasons for these limitations also suggests certain steps that might be taken to overcome these limitations and to enhance transfer and retention of trained performance.

Our most pertinent evidence documenting these limitations comes from a task that requires mental arithmetic (Rickard, 1992). Subjects were trained extensively to perform simple, single-digit mental calculations (either multiplication or division). Training was limited to the subset of problems based on operand pairs of the digits 1-9, excluding squares, in a single operand order. For example, if "12 = 3 x 4" was one of the problems selected for training, "12 = 4 x 3" was not a part of the training series. Each training set consisted of 18 multiplication problems and 18 division problems. The subject was shown all problems within a training set (constituting a block of training trials) before any problem was repeated. Forty blocks of training occurred across three sessions, the last of which also included a posttest. In the posttest, subjects were given two blocks of problems, each containing four versions of each of the training problems. One of these versions was the same as that used in acquisition (e.g., _ = 4 x 7); the three others were transformed versions, serving as tests of transfer. The manipulations used to create transfer versions of training problems were (a) a change of operand order (e.g., _ = 7 x 4), (b) a change of operation

(multiplication to division or division to multiplication; e.g., 28 = _ x 7), and (c) both operand order and operation change (e.g., 28 = _ x 4). Thus, the posttest consisted of all four versions of each problem. One month later, subjects were given a test of retention in which all four versions of each problem were presented on all four blocks of trials. The problems were presented on a CRT, and subjects typed their answers on the numeric keypad of the computer keyboard. They worked at their own pace, with a new problem appearing after feedback for the subjects' response to the preceding problem.

There are four points we wish to make about the limitations of training in this study. First, acquisition of skill during training might vary from totally specific to highly general. If training effects are general, then all problems within the same domain (single-digit operand problems) should benefit from practice on a subset of problems. On the posttest, performance should be roughly equivalent on all versions of the training problems. If transfer is specific, then only performance on trained problems should benefit from training. An intermediate position would suggest positive transfer effects to related arithmetic problems, such as problems with reversed operand order, but little or no effect to less related problems, such as those that involve a change of operation.

Our data show specific transfer of training. As shown in Figure 1, which presents results only for test problems involving multiplication, any change in problem format at the posttest had negative impact on performance. The degree of impact depended on the type of transformation made (operand order change versus operation change). But in all cases, performance was worse on transfer in contrast to training problems, suggesting that effects of training were specific to some extent to the problems used in training. We (Rickard, Mozer, & Bourne, 1992) are working on a simulation model based on interactive-activation principles which is designed to account for the present transfer results as well as interference (i.e., priming and error) patterns that have been reported elsewhere in the mental arithmetic literature (see, e.g., the recent review by Ashcraft, 1992).

Second, the posttest constituted a condition of contextual interference (or variability; see Battig, 1979, and the section below on contextual interference in acquisition of logic rules). Problems practiced during training appeared in the posttest within the sequential context of other related problems. We would expect contextual variability to have a negative impact on performance during testing (although possibly leading to better retention on some later occasion, as we will discuss again shortly). In fact these interfering effects were reflected in the data comparing the no-change problems to the end of practice, causing roughly a 50 to 60 msec drop off in average performance on problems practiced during training.

Third, performance on transfer problems provided a way to identify two processing components of the mental arithmetic task, both of which benefit from training. One of these components was more concrete or perceptually-based, corresponding to the particular digits, in all of their characteristics including order, which comprise the problem. If any change in these perceptual characteristics was made between training and transfer, performance suffered. The second component of each task was more abstract or conceptual and related to the calculation required by the problem, in this case multiplication or division. A change between training and transfer in the operation required by the problem had a more substantial negative impact on performance than did a concrete operand order change.

Finally, the impact of a one-month retention interval was more severe for the concrete, perceptible elements of the task than for the more abstract calculational elements. The only significant performance loss over the retention interval appeared in problems used in training (this effect was most salient for test problems involving division). All other problems, involving operand order change, operation change, or both, showed little loss over the one-month retention interval. Thus, just as in language-based memory, as involved, for example, in sentence comprehension (e.g., Sachs, 1967), what is lost in time from the calculation task may be primarily surface information, such as operand order. The more

abstract cognitive aspect, relating to an understanding of the material or the problem domain, may be highly resistant to the effects of disuse.

Overall, what these results suggest for training routines designed to optimize durability and transferability of training is that (a) problems used in training somehow must capture the variety of problems eventually to be encountered and (b) training should be focused on the abstract, understanding level of the task which, in contrast to more specific surface features, can be expected to be more durable over time.

III. Guidelines for Improving Long-Term Retention

With these caveats in mind, let us discuss our research on the general optimization guidelines outlined earlier, starting with the class of guidelines concerning optimization of the conditions of training.

A. Contextual Interference in Acquisition of Logic Rules

Our work on optimizing training conditions includes a project on the acquisition and retention of logic rules (Schneider, 1991). This project pursues the contextual interference effect (Battig, 1979), defined as superior memory and greater intertask transfer for materials that are particularly difficult or presented under conditions of high interference. It has been shown that varying the processing requirements from trial to trial interferes with acquisition but aids retention and transfer (see, e.g., Battig, 1979; Carlson & Yaure, 1990). Presumably, items that have more contextual interference require more processing, and are thus learned more slowly, but if well learned initially will be retained as well as, or better than, the low-interference items. This finding is of clear importance to the study of long-term skill retention because it implies that the methods used to optimize performance during acquisition are not necessarily those that will optimize performance during subsequent retention tests.

The purpose of our study was to compare practice schedules in which different procedural rules were intermixed randomly or blocked together. We used a display meant to

simulate a simplified aircraft instrument monitor consisting of four panels, only one of which was relevant (or operational) on any given trial. The relevant panels contained two lines of X's or O's in one of four combinations: XXX and XXX, XXX and OOO, OOO and XXX, OOO and OOO. The subjects' task was to decide whether or not the display in the relevant panel indicated an emergency. Each panel involved a different logical rule on which the decision was to be made. The four rules were: AND, OR, NAND, and NOR. For example, for the AND rule, an emergency was indicated only if both stimuli contained X's (i.e., XXX and XXX).

Our first experiment included one group of subjects given blocked practice (in which all trials within a block involved the same rule, i.e., the same panel, although the particular stimulus configuration varied randomly) and a second group given random practice (in which both the rule and the stimulus varied randomly from trial to trial). All subjects started with an acquisition phase followed immediately afterwards by two test blocks, one consisting of blocked rules and the other consisting of random rules.

The results of the acquisition phase in terms of correct log response time, ln(RT-200 ms), showed that the random group yielded longer response times ($\underline{M}$ = 6.691) than did the blocked group ($\underline{M}$ = 5.970), in accord with previous findings that random practice leads to strong contextual interference.

Although blocked practice led to significantly shorter response times during the acquisition phase, it led to longer latencies on the test. There was a significant interaction of practice schedule and test type, so that subjects were slowest when exposed to the blocked practice schedule and given the random test (blocked practice, random test $\underline{M}$ = 7.066; blocked practice, blocked test $\underline{M}$ = 6.192; random practice, random test $\underline{M}$ = 6.575; random practice, blocked test, $\underline{M}$ = 6.067). These findings are in accord with predictions based on contextual interference.

In our second experiment we used a third practice schedule to examine whether the unpredictability of the rules in the random group, rather than the need to retrieve the rules,

is at the heart of the contextual interference effect. This condition presented the rules in a fixed serial order (see Lee & Magill, 1983), so that the rules were predictable, but the rules changed from trial to trial, so that they had to be retrieved on each trial.

The second experiment included only a random test at the end of the acquisition phase, and this test was repeated after a delay interval, so that we could determine whether the contextual interference effect would survive, disappear, or perhaps become magnified on a retention test. The retention intervals were one week and one month.

Results from the acquisition phase showed that the blocked practice schedule yielded the shortest correct response times ($\underline{M}$ = 5.593), and the serial practice schedule ($\underline{M}$ = 5.929) yielded times midway between those of the blocked and random ($\underline{M}$ = 6.402) conditions, suggesting that both unpredictability and the need for rule retrieval contribute to contextual interference. Thus, blocked practice led to superior performance during acquisition.

In contrast, blocked practice led to inferior performance (i.e., longer response times) during both the immediate test (blocked $\underline{M}$ = 6.505, serial $\underline{M}$ = 6.367, random $\underline{M}$ = 6.204) and the long-term retention test (blocked $\underline{M}$ = 6.488, serial $\underline{M}$ = 6.457, random $\underline{M}$ = 6.310). This result was also found for proportion of correct responses (immediate test: blocked $\underline{M}$ = .876, serial $\underline{M}$ = .948, random $\underline{M}$ = .977; retention test: blocked $\underline{M}$ = .946, serial $\underline{M}$ = .949, random $\underline{M}$ = .984). Note that subjects given blocked practice made significantly fewer correct responses during the tests than subjects given random practice, even though they made more correct responses during training. Also note that there was no forgetting evident between the immediate and delayed tests; indeed accuracy improved for the blocked condition on the retention test relative to the immediate test, perhaps because the subjects got practice at rule retrieval during the immediate test, in which the rules were presented in a random order. In sum, our findings support the principle that contextual interference promotes superior performance after training. This benefit seems attributable largely to the practice

subjects received in retrieving the rules from memory. More generally, it seems crucial to match the conditions of training with the conditions required during subsequent tests.

B. Part-Whole Training in Morse Code Reception

In work on Morse code reception (Clawson, 1992), we considered the possibility that part-whole training procedures might enhance long-term retention. Specifically, we attempted to determine conditions under which independent training sessions on parts of the material would yield better acquisition and retention performance than would providing training on all the material from the beginning. In addition, we addressed a related question concerning whether any initial partial training should be restricted to the easiest material or to the most difficult material. A recently published visual discrimination study by Pellegrino, Doane, Fischer, and Alderton (1991) demonstrated that the most effective training started with the more difficult stimuli. This result could not be generalized straightforwardly to Morse code reception, of course, because Pellegrino et al.'s task was a simple visual discrimination task, whereas Morse code reception is a difficult auditory identification task. Therefore, we sought to determine whether the advantage for initially difficult training would also be found with Morse code training. Further, we were interested in whether this training advantage would also be evident on a delayed retention test.

In our first two experiments subjects learned to receive Morse code signals and to translate them to their letter equivalents. For example, subjects would hear the series of beeps short-long-short (or "di-da-di") and would be expected to respond by typing the letter "R" on a computer console. In our first experiment, subjects learned to receive 12 Morse code-letter pairs. We divided this set of pairs into two equal-sized subsets, one containing the easy items and the other containing the difficult items.

All subjects were given three sessions of training followed a month later by a retention session. During the first day of training, the subjects were divided into three groups. In the "easy-first" group, subjects received initial training on only the easy subset of code-letter

pairs; in the "difficult-first" group, subjects received initial training on only the difficult subset; whereas in the "all-first" group, subjects received training on all the letters from the beginning. After the first session, training for all subjects involved the full set of 12 code-letter pairs. During each of the four sessions, the training period was preceded by a pretest and followed by a posttest. On all days including the first, these tests covered all 12 code-letter pairs.

The results are summarized in Figure 2 in terms of proportion of correct responses on the pretests and posttests on each of the four sessions. Note that subjects in all three conditions showed similar levels of improvement across the first three days of training. However, the difference among the groups became evident immediately after the month-long retention interval, that is, on the retention pretest. Surprisingly, in light of the findings from Pellegrino et al. (1991), the difficult-first group showed a strong drop in performance at that point, whereas little forgetting was evident for the other groups.

To explore further this intriguing finding, we conducted a second experiment that included only the easy-first and difficult-first training groups with a substantially greater number of subjects in each group. Further, we altered our procedures to facilitate the recording of response times.

As shown in Figure 3, which summarizes the accuracy results broken down by the two types of letter pairs (easy and difficult), we found once again a larger drop in performance on the retention pretest for the difficult-first group than for the easy-first group, but in this case the difference between training groups was only found on the easy pairs.

Figure 4 shows the results in terms of mean correct response time, rather than accuracy. As for proportion of correct responses, we found worse performance (in this case, slower responding) for the difficult-first group than for the easy-first group on the easy pairs in the retention pretest. However, on the difficult pairs in that test, the difficult-first group was faster than the easy-first group. Despite this one advantage for the difficult-first

group, in general, performance after training was inferior in our study when the difficult items were studied first. This finding contrasts to that of Pellegrino et al. (1991) who found that difficult-first training was superior. However, as we noted previously, there were many differences between our investigation of Morse code and Pellegrino et al.'s investigation of visual discrimination. We think the most crucial difference concerned the definition of the easy items. In our task the easy items were in fact quite challenging, performance on them being near 50% accuracy initially. In Pellegrino et al.'s task the easy items were truly easy, performance being at the ceiling during the initial training phase. When subjects must devote their initial training to very easy items, it is not surprising that they do not develop the strategies that would help them with more challenging material presented later.

The aim of our third experiment then was to localize the sources of difficulty for all the Morse code stimuli. Toward this end, we divided the Morse reception task into parts, not in terms of different stimuli to be learned, but rather in terms of subtasks to be performed.

All subjects in this experiment studied all 12 of the stimulus letters simultaneously, but there were three groups who studied them differently. The code-to-letter group was trained in the normal reception task of hearing the codes and typing their corresponding letters; the code-to-dida group heard the codes and typed keys corresponding to "di" (short) and "da" (long), segmenting the auditory code into its elements; and the dida-to-letter group read simplified di-da patterns displayed on the CRT and translated the segmented signals into their corresponding letters, which they typed on the keyboard. In this experiment subjects were trained for two sessions with a pretest at the start of training, a posttest at the end of training, and a retention test two weeks later.

The results are summarized in Figure 5 in terms of proportion of correct responses. Accuracy for the code-to-dida group was remarkably stable, showing only small (but significant) improvement as training progressed, whereas accuracy for the dida-to-letter and code-to-letter groups improved considerably across the acquisition sessions. Also, there was

some forgetting across the two-week retention interval for the dida-to-letter and code-to-letter groups, but no forgetting for the code-to-dida group. The code-to-dida task involved a skill that was largely based on perceptual procedures, whereas both of the other tasks required the learning of paired associates. Our finding of no forgetting in the code-to-dida group but substantial forgetting in the other two groups is consistent with our previous observation that memory based on procedures, in contrast to memory for facts (or verbal associations), is highly resistant to forgetting over long delays (Healy et al., 1990, 1992).

Analyses of individual differences suggested that the code-to-dida group was more stable than were the other two groups across the three sessions. For subjects in the code-to-dida group, accuracy on the posttest and retention test was predictable from pretest scores, whereas for the other groups the correlations were all nonsignificant. This finding of stability for the code-to-dida group suggests that the processes involved in segmenting the auditory signal into elements may be the limiting factor leading to the failure of some students of Morse code to learn the reception task successfully. That is, individuals who are poor at the code-to-dida task may not be able to improve performance on the full code-to-letter task even with much practice.

In a post hoc analysis, we examined the extent to which separate performance on the component subtasks could predict performance on the whole Morse code reception task. For this analysis we computed a predicted accuracy level for the whole (code-to-letter) task based on the product of the observed accuracy levels for the two part tasks. Although there was no difference between observed and predicted whole task performance at the pretest (observed $\underline{M}$ = .304, predicted $\underline{M}$ = .313), observed whole task performance tended (nonsignificantly) to exceed predictions at both the posttest (observed $\underline{M}$ = .756, predicted $\underline{M}$ = .644) and the retention test (observed $\underline{M}$ = .661, predicted $\underline{M}$ = .576), suggesting that subjects may develop effective strategies in the whole task to overcome problems encountered in the partial component tasks.

C. Part-Whole Training of Tank Gunner Skills

Whether acquisition and retention benefit from part-whole training was also a focus of our research with tank gunner skills (Marmie & Healy, 1992). Like the first Morse code study, we examined whether there was superior transfer to the whole task from part- or whole-task training. Like the last Morse code study, we have broken down the whole task into sequential component subtasks.

In this study subjects were engaged in a realistic, goal-directed simulation exercise. The advantage of using this simulation exercise in part-whole training was threefold: First, it was a task which subjects generally found intrinsically motivating because of its similarity to an arcade video game. In contrast, for example, the important tests of part-whole training by Naylor and Briggs (1963) used training on a laboratory Markov prediction task which seems less intrinsically motivating. Second, our division yielded clearly separable, meaningful, goal-directed subtasks (see Newell, Carlton, Fisher, & Rutter, 1989, who also recommended the use of natural subtasks). In contrast, for example, in a more recent study of part-whole training with a video game environment, Mane, Adams, and Donchin (1989) found it necessary to have subtasks be repetitive drills. Third, and most important, the simulation exercises we used had separate dependent measures that allowed us to examine the specific decay of task components over a retention interval.

More specifically, in our study, stimuli were presented on TopGun Tank Simulators. The simulators utilized color monitors mounted in an enclosed sit-down unit, which was designed as a training machine for tank gunners. Subjects in our experiment controlled tank gun turret movements via hand controls and aimed at threat targets with the aid of a sight. Two digitized human voices played the roles of the commander and the loader, telling the subjects where to lay on their sight, when they had ammunition loaded and available for use, and when to fire. A schematic display of a target tank, as viewed by subjects looking at the simulator monitor, is presented in Figure 6. Each session included a presentation of 100

target tanks divided into 10 blocks of 10 trials each, with each tank shown for a maximum of 20 seconds.

The subjects' tank could not move, but the hand controls which moved their sight allowed them 360 degree visibility. Subjects fired by pressing either of two buttons under their index fingers. A threat tank was destroyed when a shot struck its center of mass. The result was scored as a "kill." We tabulated kills and two different response-time measures: time to make an identification and time to fire (after an identification had been made). A tank was considered identified when it entered the subject's field of view. The identification, or ID, measure reflected the search component of the task, or how long it took the subject to find the target. The time-to-fire measure reflected the combined subsequent components of sighting (or laying on the sight) and firing (or shooting at the tank). In general this measure reflected development of the sighting skill, which was the most difficult of the three components. Both of these measures were computed only for successful kills of the target tank.

Subjects were tested over four sessions. The first three sessions occurred during a single week, with the last session occurring four weeks later. The experiment employed two groups of subjects. During the first two sessions, the part-training group engaged in part-task training, practicing the sighting and firing task subcomponents (which are indexed by the time-to-fire measure). But training was not given on the search component (which is indexed by the ID measure); the simulators were programmed, using an optional function called "autoslew," to relieve the gunner of the requirement to ID the enemy threat (as occurs in a real tank when the commander assumes control of the ID task). The last two sessions involved the whole task, combining sighting and firing with searching. The whole-training group engaged in whole-task training and was trained on all three subcomponents of the task simultaneously throughout all four sessions.

Performance on the search component of the task is summarized in Figure 7 in terms of mean time to ID in seconds. Note that because of the autoslew function, the first two

sessions for the part-training group reflect the performance of the simulated commander, not that of the subjects. Although during the first two sessions the part-training subjects received no practice on the search component of the task, after the initial training they performed just as well as the whole training group. Figure 8 shows mean time to fire. Note that the subjects given part training, which may have allowed them to concentrate on the sighting and firing subcomponents of the task initially, showed a large advantage in the second session of training, and that advantage was maintained after initial training, even during the retention test. Holding back on the training of the search subcomponent benefited the sighting and firing subcomponents with undistracted practice, and that benefit persisted even after the search subcomponent was introduced into the task.

Although it benefited response time performance, part training did not appear to aid subjects in improving their accuracy on the task. Accuracy results are summarized in Figure 9 in terms of mean proportion of kills. Note that during initial training, there was a substantial advantage for the part training group because the commander efficiently took over the searching component of the task. After the initial training period, there was no difference between the two training groups.

The combined findings across the three different measures of performance indicate that part training does not hurt performance relative to whole training, and may in fact improve performance by allowing subjects to concentrate on one task at a time. By comparing this finding to that obtained in our initial Morse code experiment, which found a clear disadvantage for initial training on the difficult subcomponent, it is clear that any conclusions concerning part-whole training depend crucially on the nature of the whole task and the characteristics of the component part tasks. In particular, we attribute the disadvantage for the difficult-first condition in the Morse code study and the contrasting advantage for the (difficult-first) part-training condition in the tank gunner study to the fact that the difficult items in the Morse code reception task could not be mastered within the time allotted, whereas the sighting and firing

components of the tank gunner task could be mastered during the initial training period. More generally, when training on only a part of a whole task, it seems crucial to focus on a component that is sufficiently complex to be engaging but not so complex to be impossible to master in the time allowed.

D. A Generation Advantage for Mental Arithmetic and Vocabulary Learning

We have discussed some ways to manipulate the conditions of training in order to optimize long-term retention, including blocked/random and part/whole comparisons. Another powerful manipulation of training conditions is the comparison of reading and generating. The generation effect (see, e.g., Slamecka & Graf, 1978) refers to the finding that people show better retention of learned material when it is self-produced, or generated, than when it is simply copied, or read. The typical task used to investigate the generation effect has been one in which the subject is presented a series of paired associates in either a read or a generate format and is subsequently required to recall or recognize the second item of each pair. Thus, the subject's task is to recall the occurrence of a prior *event* or *episode*, in this case the prior occurrence of a paired associate. Previous studies of the generation effect have been limited almost exclusively to examinations of memory for episodes or events (see, e.g., Crutcher & Healy, 1989). In contrast, our recent work (McNamara & Healy, 1991) extended this finding to memory for facts and skills, including multiplication skill. In accordance with our procedural reinstatement framework (see Healy et al., 1992), we proposed that a critical factor leading to a generation advantage for skill training is that stable and efficient cognitive strategies be developed during the training process. Multiplication is a skill for which most college students have already developed some cognitive strategies. For simple single-digit operand problems, we would expect no change in these strategies as a function of training because they are extremely well entrenched. In fact, answer retrieval might be or become automatic (see the section below on direct and mediated retrieval in mental arithmetic). In contrast, most college students have not developed stable cognitive

strategies for more difficult multiplication problems with operands greater than 12. Thus, only for these difficult problems would a generation advantage be expected because the generate condition would be more apt than the read condition to promote the formation of new cognitive strategies.

We tested this prediction by comparing read and generate conditions of training on both easy (e.g., 40 x 9 = 360) and hard (e.g., 14 x 9 = 126) multiplication problems. Subjects were given a pretest, training in either the read or generate condition, and a posttest, all on multiplication problems. In the read condition, subjects were presented the multiplication problem and answer on the computer screen; for example, "40 x 9 = 360". They copied the problem and answer by typing them on the number pad. In the generate condition, subjects were presented the problem on the computer screen; for example, "40 x 9 = ". They then typed the problem and the answer that they generated.

The results of this study are summarized in Figure 10 for proportion of correct responses. In accord with predictions, a generation advantage was found only on the hard problems in the posttest.

The arithmetic material studied in this experiment was already familiar to the subjects before training. Of great interest would be the extension of this investigation to situations in which individuals are learning new material. Such a question has important implications for the many training situations that involve teaching new material, rather than improving the efficiency with which old material is retrieved.

Therefore, in our next experiment we had subjects learn word-nonword associations. The findings from our last experiment suggested that the use of cognitive strategies aids learning and retention. However, we did not directly assess strategy use in that study. Hence, in the present experiment we directly examined the strategies used by the subjects. The most probable relevant cognitive strategies in this case were mnemonic codes linking the word and nonword components of each pair. We expected subjects in the generate condition to develop

more mnemonic codes than subjects in the read condition and, therefore, to show superior learning and retention of the word-nonword pairs. We further expected that subjects in the read condition who developed mnemonic codes would show a level of performance comparable to that of subjects in the generate condition. To assess the extent of mnemonic coding, we administered a retrospective questionnaire asking the subjects to report their use of mnemonic codes for each word-nonword pair.

Subjects were given a list of 30 word-nonword pairs to study for ten minutes before training began. They were then administered a pretest, followed by training, and then a posttest. To evaluate the long-term impact of both training and mnemonic coding, we included a retention test after a one-week delay.

As expected, the generation advantage was only evident after training; that is, on both the posttest and the retention test. Specifically, on the pretest there was no advantage in terms of the proportion of correct responses for the generate condition ($\underline{M}$ = .297) relative to the read condition ($\underline{M}$ = .353), whereas on the posttest the generate condition ($\underline{M}$ = .956) was superior to the read condition ($\underline{M}$ = .833). Likewise, on the retention test, the generate condition ($\underline{M}$ = .756) showed higher accuracy than the read condition ($\underline{M}$ = .658).

In order to pinpoint the locus of the generation advantage, we categorized the subjects in each training condition into those with a relatively high and those with a relatively low average mnemonic score on the basis of the retrospective questionnaire. Figure 11 presents the proportions of correct responses separately for low and high mnemonic subjects. As predicted, subjects in the read condition who used mnemonic coding showed a level of performance on the posttest and retention test comparable to that shown by subjects in the generate condition.

Did the likelihood of recalling a particular nonword depend on whether subjects employed a mnemonic strategy to encode it? Figure 12 shows that the overall proportion of correct responses was highest for the items given high mnemonic scores and lowest for the

items with no mnemonics. Crucially, forgetting across the retention interval was least for items given high mnemonic scores. This finding suggests that a mnemonic strategy aids not only coding but also long-term retention of information. More generally, this finding indicates that to maximize long-term retention it is crucial to optimize not only the conditions of training but also the learning strategy used by the subjects.

E. Direct and Mediated Retrieval in Mental Arithmetic

Facts can be retrieved from memory in one of two ways, either automatically (by direct access to a fact network) or indirectly (by some mediated route). In the latter case, retrieval is deliberate, conscious, and effortful, whereas in the former case it occurs effortlessly.

Direct access is not a characteristic of tasks but rather of facts or skill components of a task. Within any particular task domain, direct access co-exists with mediated retrieval. For example, we have shown in a mental arithmetic task that sometimes answers are achieved directly and other times indirectly (Bourne & Rickard, 1991). Note that mental arithmetic is a skill that most adults will claim already to have. Here we are interested in the effects that further practice has on a known skill. As we will see, performance is based partly on direct and partly on mediated answer retrieval, and a transition from indirect to direct retrieval may be an important consequence of further training that might have major implications for long-term memory.

In one study, we gave subjects two one-hour sessions of practice on 25 selected single-digit multiplication problems. These problems were presented to subjects one at a time in blocks. Each of the two sessions consisted of 30 blocks of 25 problems each. In the first two blocks of each session subjects were asked, after responding to each problem, whether the answer popped into mind directly or had to be retrieved through one or more consciously-mediated steps. An example of mediated performance is based on an anchor-and-

adjust strategy: Asked to provide an answer to "8x6," the subject retrieves 8x5=40 (anchor) and adds 8 (adjust).

About 18% of the problems in the first two blocks were solved by mediation. There was some variability among subjects, who ranged from no mediation (all direct retrievals, by self-report) to about 60% mediation. We observed both intrasubject and intraproblem stability in these data. That is, if a subject reported mediation on Block 1 of Session 1, he or she was also likely to report mediation on Block 2 of Session 1. Likewise, if a particular problem was mediated on Block 1, it was likely to be mediated on Block 2 as well. It will come as no surprise that, when mediation was reported, the subject was slower to respond with the correct answer. Figure 13 shows response time on the first two blocks of Session 1 for problems that were mediated on both occasions ("Both Other"), on only one occasion ("Other 1" and "Other 2"), or on neither occasion ("Both Direct"). The data of subjects who never reported mediation ("All Direct") are also included for comparison. Somewhat more interesting is the fact that the effects of mediation persisted throughout the entire experiment. In Figure 14, we show response time on all sixty blocks of practice (Session 1 and Session 2) for problems identified as direct or other on the first two blocks of Session 1. We interpret the fact that response time differences persisted to suggest that, if a problem was mediated early in the training session, it had a high probability of continuing to require mediation throughout the remaining blocks of training. Supporting this argument are the data from Blocks 1 and 2 from Session 2. Approximately the same number of problems (16%) required mediation on the second session as on the first session. Moreover, there was a strong correspondence between subjects reporting mediation and between problems requiring mediation in the two sessions. Although subjects became faster with training, it, thus, does appear that the method by which a given subject solved a given problem remained stable over a large number of repetitions. This finding may pose a challenge for Logan's (1988) influential instance theory of automatization which suggests that increased learning leads to a

transition from mediated to direct retrieval. Further investigations with more extensive practice are needed to resolve this issue.

F. Direct and Mediated Retrieval in Vocabulary Acquisition

In order to study the transition from mediated to direct retrieval under controlled practice conditions, it may be preferable to study the acquisition of new knowledge and then study changes in retrieval as a function of extended practice. A particularly attractive task domain to study retrieval of new knowledge is the learning of vocabulary items in a foreign language. We have in several studies instructed subjects to learn Spanish vocabulary items with the keyword method (see, e.g., Crutcher, 1990, 1992; Healy et al., 1992).

In the keyword method, the Spanish word (e.g., doronico) is first related to a keyword, a concrete English word similar in sound to the Spanish word (e.g., door). The keyword is then associated to the English equivalent (leopard) by forming an interactive image (e.g., a leopard walking through a door). This method of learning provides a great deal of control over mediational processes, thus assuring a similar encoding structure across all subjects.

We have shown that retrieval of English equivalents after original acquisition was virtually always mediated by retrieval of the keyword in working memory, based on three sources of information (Crutcher, 1992).

First, the retrieval times for the English equivalent of the Spanish word, the Vocabulary Task, were substantially slower (M = 2,041 ms) than those for the two subtasks, the Keyword Subtask (M = 1,653 ms), which involves responding with the similar-sounding English keyword given the Spanish word as a cue, and the English Subtask (M = 1,633 ms), which involves responding with the English translation given the keyword as a cue.

Second, retrieval accuracy for the English equivalent of the Spanish word after a delay of a week, a month, or a year was a direct function of accuracy on the two subtasks. That is, accurate retrieval of the English equivalent was virtually only observed when both subtasks were accurately performed at the retention test.

Third, retrospective verbal reports after successful retrievals revealed that subjects reported accessing the keyword prior to accessing the English equivalent. Indeed, when subjects reported retrieving the English equivalent directly, the retrieval time was over 500 milliseconds faster. Hence, we can conclude that retrieval after original learning was mediated.

We have also studied the effects on retrieval of 80 additional retrieval trials with each item in several sessions spread out over two weeks. After initial acquisition and test, subjects practiced the Vocabulary Task (full practice) for half of the items and the subtask of retrieving the English equivalent using the keyword (subtask practice) for the other half of the items. The retrieval times for the initial test and the final test after extended practice are shown in Figure 15; note that the tests included both tasks ("Vocabulary Task" and "English Subtask") for all items.

At initial test the Vocabulary Task was reliably slower than the English Subtask, which replicated the earlier finding of mediated retrieval following acquisition. At the final test after practice, a reliable cross-over interaction was found, in which the items in the full practice condition were retrieved faster with the Vocabulary Task than with the English Subtask, with the opposite result for items in the subtask-practice condition. An analysis of retrospective reports for only the Vocabulary Task revealed that at the initial test, subjects reported retrievals involving the mediation of the keyword for both the subtask-practiced items ($\underline{M}$ = 86.7%) and the full-practiced items ($\underline{M}$ = 83.8%), but at final test most retrievals for full-practiced items involved no reported mediation ($\underline{M}$ = 15.6%), although most retrievals for subtask-practiced items continued to involve reported mediation ($\underline{M}$ = 87.0%). We are currently analyzing data from a one-month retention test of these subjects.

These results clearly showed that after extended practice retrieval was no longer a sequential process involving access of the keyword in working memory. One possibility is that a genuinely different association was formed between the Spanish word and its English

equivalent. Another possibility is that retrieval still involved the keyword, but with extended practice access involved covert mediation of the keyword through spreading activation. In support of the latter interpretation we showed in a new experiment that learning a new association to an old keyword interfered with subsequent retrieval of the original Spanish-English pair, even when that original pair had been extensively practiced. In agreement with an earlier study on the effects of extensive practice (Pirolli & Anderson, 1985; see also our own research with mental multiplication, Bourne & Rickard, 1991), we demonstrated in this study that the original encodings with their mediators continued to exert their influence after extensive practice even when the observable characteristics of the retrieval process suggested direct retrieval.

G. Automatic Processing in Color-Word Interference

Much of our initial work on the long-term retention of skills (see, e.g., Healy et al., 1992) was guided by a hypothesis relating superior retention, or entry into permastore, to the achievement of automatic processing, or direct retrieval, during acquisition. We attempted to test this hypothesis in two different domains, the first involving target detection and the second involving mental multiplication (see Healy et al., 1990, 1992; Fendrich et al., in press). We did find superior long-term retention in both of those studies, but we have as yet been unable to establish conclusively that automaticity was achieved by our subjects and, therefore, whether there was a clear relationship between automatic processing and long-term retention. We propose that the task that might hold the key to resolving this issue is the familiar Stroop color-word interference task. In the Stroop task, subjects are asked to name the color of the ink in which color words are printed. The ink color and word do not correspond. For example, given the word purple printed in red ink, the appropriate response is "red." This task has been widely accepted as demonstrating that word reading is automatic and hence interferes with the nonautomatic task of color naming (see, e.g., MacLeod, 1991, for a recent review of research on the Stroop effect). Our proposed study involves the

training of the color naming task to the point of automaticity so that no interference would be evident. Some of us (Clawson, King, Healy, Ericsson, & Marmie) are training subjects in two different color-naming situations. The first training condition involves practice in simply naming color patches. The second training condition involves practice in naming the colors of incongruent color words. Examining the effects of training on performance in the Stroop task should enable us to resolve the issue of interest concerning long-term retention, and should also allow us to disentangle the competing theories that have been proposed as explanations for the Stroop effect (see MacLeod, 1991).

The results from two pilot subjects are shown in Figure 16. One subject was given training only on the color patches task (shown on the left of Figure 16) and the second subject was given training only on the Stroop task itself (shown on the right of Figure 16). These preliminary subjects were given a pretest on one day, a single hour of training the next day, and then a posttest the following day. The pretests and posttests included both color patch naming and Stroop tests. Note that despite the fact that these pilot subjects were given only one hour of training, as opposed to the 12 hours of practice planned for the full experiment, we found substantial decreases in response times from the pretest to the posttest on both tasks for both subjects, suggesting that indeed training will prove to have profound effects and lead to automatic color naming responses. Hence, we are encouraged that this study will allow us to elucidate the relationship between automatic processing and long-term retention.

IV. Summary and Conclusions

In closing, we will review the three classes of guidelines we found to optimize long-term retention (see Table 1 for a summary). The first class of guidelines concerned ways to optimize the conditions of training. We discussed three general guidelines in this class. The first concerned the contextual interference found, for example, with random sequences of tasks as opposed to fixed or predictable sequences. Although random sequences did suppress performance during acquisition, they promoted superior performance after training. We

attribute this benefit in large part to the practice subjects received in retrieval from memory of the appropriate response preparation procedures and to the necessity to match the conditions of training with the characteristics of the desired target performance. The second general guideline in this class concerned training parts of a task versus the whole task. We conclude from our findings that it is best to focus initially on a maximally trainable component of the task, that is, to avoid wasting time on either a trivial component or a component that cannot be adequately mastered within the constraints of the training period. The third general guideline in this class concerned the distinction between generating and reading. We conclude that the well known generation advantage can be extended from memory for episodes to memory for facts and skills.

The second class of guidelines concerned ways to optimize the strategies used. We found that in tasks that require deliberate retrieval from memory, training that promotes efficient encoding strategies maximizes long-term retention.

The third class of guidelines concerned ways to attain direct access, or automatic retrieval, from memory. We found in several domains that achieving automaticity requires extensive practice. It is surprising that even after vast amounts of practice there is still mediated retrieval for a small subset of items. Further, even when retrieval appears automatic after extensive practice, mediators may still continue to exert their influence. Finally, we are in a position now to test our original hypothesis that there is a unique retention advantage for items that have achieved the status of automatic retrieval.

We started this chapter by summarizing some of our work demonstrating the specificity of improvement in performance. That is, training on specific items showed little or no transfer to related items. The specific characteristics of the training context seemed to have a profound influence on immediate transfer to other contexts. Hence, we now need to focus more on the transfer of training to new skills and on optimizing the generalizability of training. Our task had been to examine the optimization of long-term retention, but we have

learned that optimizing retention does not guarantee generalizability, and it is even possible that there is a trade-off between durability and generalizability. Our horizons have, thus, now broadened, so what we intend for the future is to explore conditions of training and strategy utilization that will simultaneously maximize both generalizability and long-term retention.

Acknowledgments

This research was supported by Army Research Institute Contract MDA903-90-K-0066 to the Institute of Cognitive Science at the University of Colorado. Robert J. Crutcher is now at the University of Illinois, Chicago; K. Anders Ericsson is now at Florida State University, Tallahassee. We are grateful to Douglas Medin and Evan Heit for helpful comments concerning an earlier version of this chapter.

References

Ashcraft, M.H. (1992). Cognitive arithmetic: A review of data and theory. Cognition, **44**, 75-106.

Bahrick, H.P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. Journal of Experimental Psychology: General, **113**, 1-29.

Battig, W.F. (1979). The flexibility of human memory. In L.S. Cermak & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 23-44). Hillsdale, NJ: Erlbaum.

Bourne, L.E., Jr., & Rickard, T.C. (July, 1991). Mental calculation: The development of a cognitive skill. Interamerican Congress of Psychology. San Jose, Costa Rica.

Carlson, R.A., & Yaure, R.G. (1990). Practice schedules and the use of component skills in problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, **16**, 484-496.

Clawson, D.M. (1992). Acquisition and retention of Morse code reception skills. Unpublished master's thesis, University of Colorado, Boulder.

Crutcher, R. J. (1990). The role of mediation in knowledge acquisition and retention: Learning foreign vocabulary using the keyword method (Tech. Rep. No. 90-10). Boulder, CO: University of Colorado, Institute of Cognitive Science.

Crutcher, R.J. (1992). The effects of practice on retrieval of foreign vocabulary using the keyword method. Unpublished doctoral dissertation, University of Colorado, Boulder.

Crutcher, R.J., & Healy, A.F. (1989). Cognitive operations and the generation effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, **15**, 669-675.

Farr, M.J. (1987). The long-term retention of knowledge and skills: A cognitive and instructional perspective. New York: Springer-Verlag.

Fendrich, D.W., Healy, A.F., & Bourne, L.E., Jr. (1991). Long-term repetition effects for motoric and perceptual procedures. Journal of Experimental Psychology: Learning, Memory, and Cognition, **17**, 137-151.

Fendrich, D.W., Healy, A.F., & Bourne, L.E., Jr. (in press). Mental arithmetic: Training and retention of multiplication skill. In C. Izawa (Ed.), Cognitive psychology applied. Hillsdale, New Jersey: Erlbaum.

Gick, M.L., & Holyoak, K.J. (1987). The cognitive basis of knowledge transfer. In S.M. Cormier & J.D. Hagman (Eds.), Transfer of training: Contemporary research and applications (pp. 9-46). New York: Academic Press.

Healy, A.F., Fendrich, D.W., Crutcher, R.J., Wittman, W.T., Gesi, A.T., Ericsson, K.A., & Bourne, L.E., Jr. (1992). The long-term retention of skills. In A.F. Healy, S.M.

Kosslyn, & R.M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes, Volume 2 (pp. 87-118). Hillsdale, NJ: Erlbaum.

Healy, A.F., Fendrich, D.W., & Proctor, J.D. (1990). Acquisition and retention of a letter-detection skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, **16**, 270-281.

Lee, T.D., & Magill, R.A. (1983). The locus of contextual interference in motor-skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, **9**, 730-746.

Logan, G.D. (1988). Toward an instance theory of automatization. Psychological Review, **95**, 492-527.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, **109**, 163-203.

Mane, A.M., Adams, J.A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. Acta Psychologica, **71**, 179-196.

Marmie, W.R., & Healy, A.F. (1992). The long-term retention of a complex skill: Part-whole training of tank gunner simulation exercises. Unpublished manuscript.

McNamara, D.S., & Healy, A.F. (November, 1991). A generation advantage for multiplication skill and nonword vocabulary acquisition. Paper presented at the 32nd Annual Meeting of the Psychonomic Society, San Francisco, California.

Naylor, J.C., & Briggs, G.E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. Journal of Experimental Psychology, **65**, 217-224.

Newell, K.M., Carlton, M.J., Fisher, A.T., & Rutter, B.G. (1989). Whole-part training strategies for learning the response dynamics of microprocessor driven simulators. Acta Psychologica, **71**, 197-216.

Pellegrino, J.W., Doane, S.M., Fischer, S.C., & Alderton, D. (1991). Stimulus complexity effects in visual comparisons: The effects of practice and learning context. Journal of Experimental Psychology: Human Perception and Performance, **17**, 781-791.

Pirolli, P.L., & Anderson, J.R. (1985). The role of practice in fact retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, **11**, 136-153.

Rickard, T.C. (1992). Acquisition and transfer of mental arithmetic skill. Unpublished master's thesis, University of Colorado, Boulder.

Rickard, T.C., Mozer, M.C., & Bourne, L.E., Jr. (1992). An interactive activation model of arithmetic fact retrieval (Tech. Rep. No. 92-15). Boulder, CO: University of Colorado, Institute of Cognitive Science.

Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception & Psychophysics, **2**, 437-442.

Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, **3**, 207-217.

Schneider, V.I. (1991). The effects of contextual interference on the acquisition and retention of logic rules. Unpublished doctoral dissertation, University of Colorado, Boulder.

Slamecka, N.J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, **4**, 592-604.

Underwood, B.J. (1964). Degree of learning and the measurement of forgetting. Journal of Verbal Learning and Verbal Behavior, **3**, 112-129.

Wisher, R.A., Sabol, M.A., Sukenik, H.K., & Kern, R.P. (1991). Individual ready reserve (IRR) call-up: Skill decay (Research Report No. 1595). Alexandria, VA: U.S. Army Research Institute.

Table 1

Guidelines for Optimizing Long-Term Retention (with Relevant Chapter Sections)

---

1. Optimize conditions of training
   - A. Promote contextual interference (Acquisition of Logic Rules, Section III A)
   - B. Focus initially on maximally trainable component (Morse Code Reception, Section III B; Tank Gunner Skills, Section III C)
   - C. Encourage generation during practice (Mental Arithmetic and Vocabulary Learning, Section III D)
2. Optimize the learning strategy used (Vocabulary Learning, Section III D)
3. Achieving automaticity is difficult but may have a unique retention advantage (Mental Arithmetic, Section III E; Vocabulary Acquisition, Section III F; Color-Word Interference, Section III G)

Figure Legends

Figure 1. Results of experiment by Rickard (1992) for test problems involving multiplication. Mean correct response time in ms as a function of test time, block, and problem version. The mean for the last block of practice is also shown for comparison. (All means were calculated based on log RTs and then transformed back to ms by the anti-log function.)

Figure 2. Results of Experiment 1 by Clawson (1992). Mean proportion of correct responses for pretests (pre) and posttests (post) as a function of initial training group and session.

Figure 3. Results of Experiment 2 by Clawson (1992). Mean proportion of correct responses on easy and difficult (diff.) code-letter pairs for pretests (pre) and posttests (post) in the difficult-first (D-1st) and easy-first (E-1st) training groups as a function of session.

Figure 4. Results of Experiment 2 by Clawson (1992). Mean correct response time in ms on easy and difficult (diff.) code-letter pairs for pretests (pre) and posttests (post) in the difficult-first (D-1st) and easy-first (E-1st) training groups as a function of session. (All means were calculated based on log RTs and then transformed back to ms by the anti-log function.)

Figure 5. Results of Experiment 3 by Clawson (1992). Mean proportion of correct responses for the three task groups on the pretest (Pre), posttest (Post), and retention test (Ret).

Figure 6. Schematic display of a target tank on the TopGun simulator monitor in the experiment by Marmie and Healy (1992).

Figure 7. Results of the experiment by Marmie and Healy (1992). Mean time to ID target in s for successfully killed targets as a function of training group and session. (Session 4 is the retention session.)

Figure 8. Results of the experiment by Marmie and Healy (1992). Mean time to fire in s for successfully killed targets as a function of training group and session. (Session 4 is the retention session.)

Figure 9. Results of the experiment by Marmie and Healy (1992). Mean proportion of kills as a function of training group and session. (Session 4 is the retention session.)

Figure 10. Results of Experiment 1 by McNamara and Healy (1991). Mean proportion of correct responses on the pretest and posttest as a function of training condition and problem difficulty.

Figure 11. Results of Experiment 2 by McNamara and Healy (1991). Mean proportion of correct responses for the subjects with low and high average mnemonic scores on the pretest, posttest, and retention test as a function of training condition.

Figure 12. Results of Experiment 2 by McNamara and Healy (1991). Mean proportion of correct responses for the items given no mnemonic, a low mnemonic, or a high mnemonic score on the pretest, posttest, and retention test.

Figure 13. Results of experiment by Bourne and Rickard (1991). Mean correct response time in ms on the first two blocks of Session 1 as a function of problem mediation. (All means were calculated based on log RTs and then transformed back to ms by the anti-log function.)

Figure 14. Results of experiment by Bourne and Rickard (1991). Mean correct response time in log ms for all 60 blocks as a function of log block and problem mediation on the first two blocks of Session 1.

Figure 15. Results of experiment by Crutcher (1992). Mean correct response time in ms for the Vocabulary task and the English subtask on the initial and final tests as a function of practice condition. (All means were calculated based on log RTs and then transformed back to ms by the anti-log function.) Standard errors of the mean shown as bars.

Figure 16. Results of pilot experiment by Clawson, King, Healy, Ericsson, and Marmie. Mean correct response time in ms for the patches and Stroop tests on the pretest and posttest as a function of training condition.

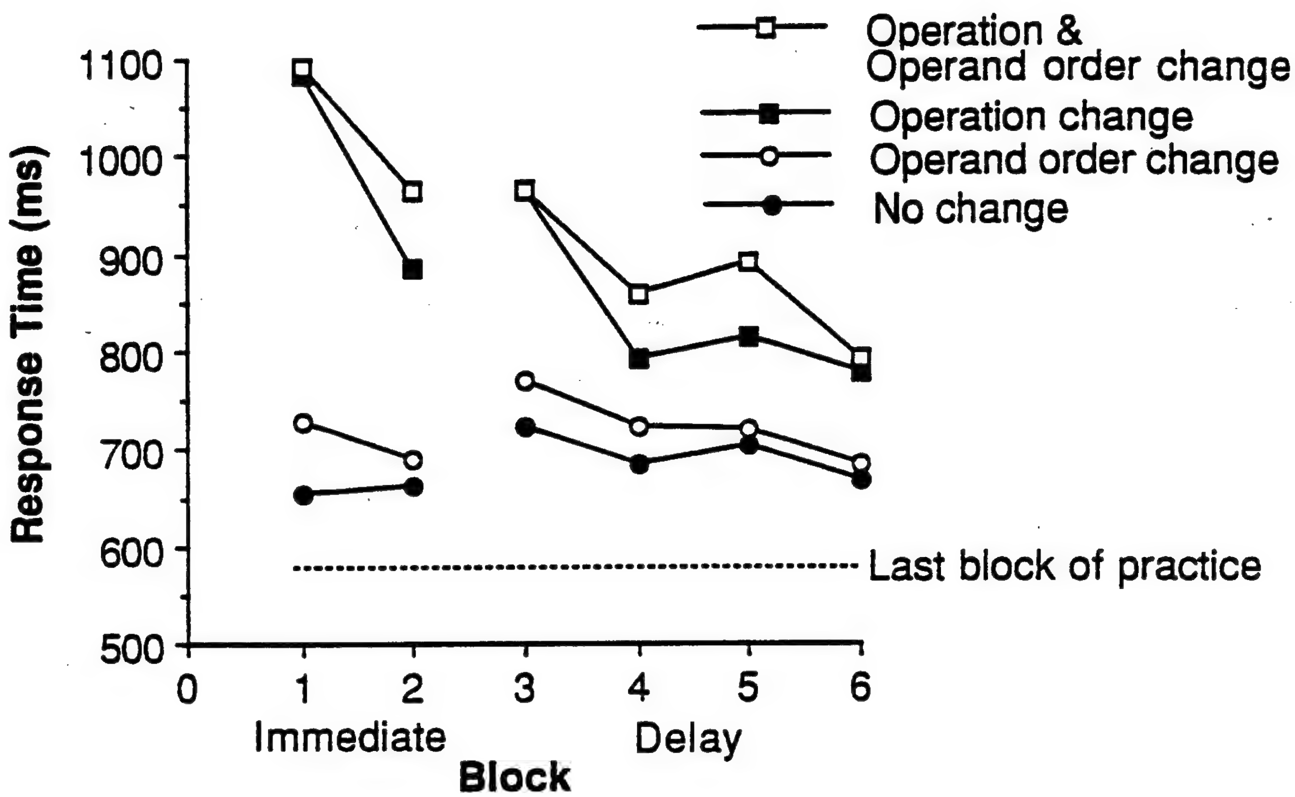
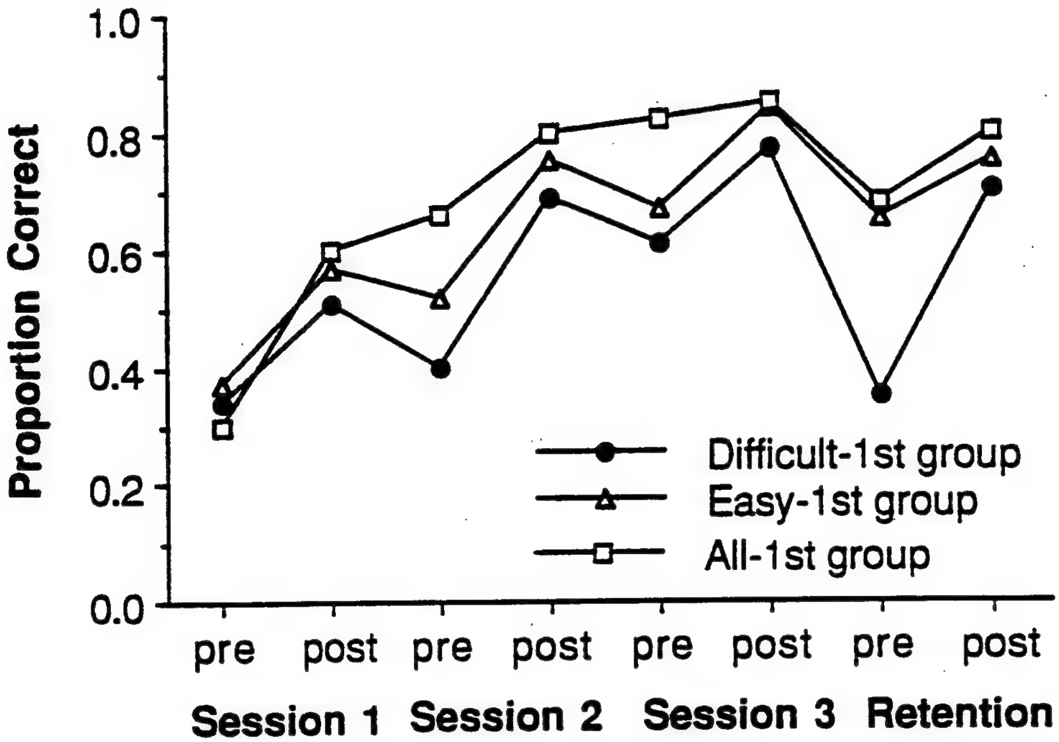

Operation &
Operand order change
Operation change
Operand order change
No change
1100
1000
900
800
700
600
500
Response Time (ms)
Last block of practice
0
1
2
3
4
5
6
Immediate
Delay
Block

Figure 1



1.0
0.8
0.6
0.4
0.2
0.0
Proportion Correct
Difficult-1st group
Easy-1st group
All-1st group
pre
post
pre
post
pre
post
pre
post
Session 1
Session 2
Session 3
Retention

Figure 2

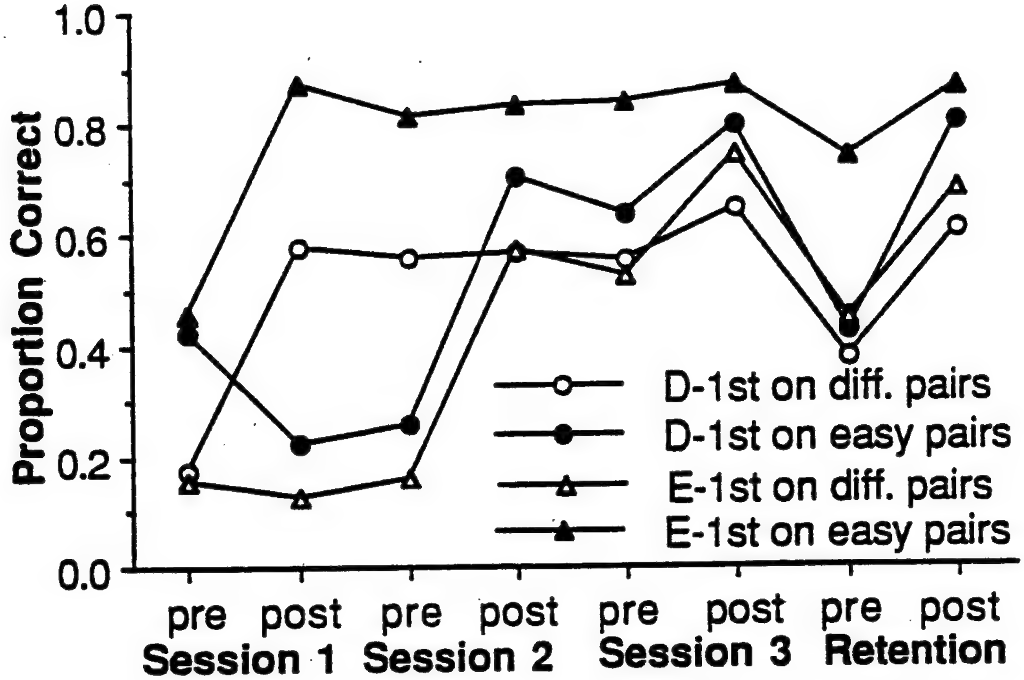


Figure3

Healy et al., The Long-Term Retention..., <u>The Psychology of Learning and Motivation</u>, Figure #

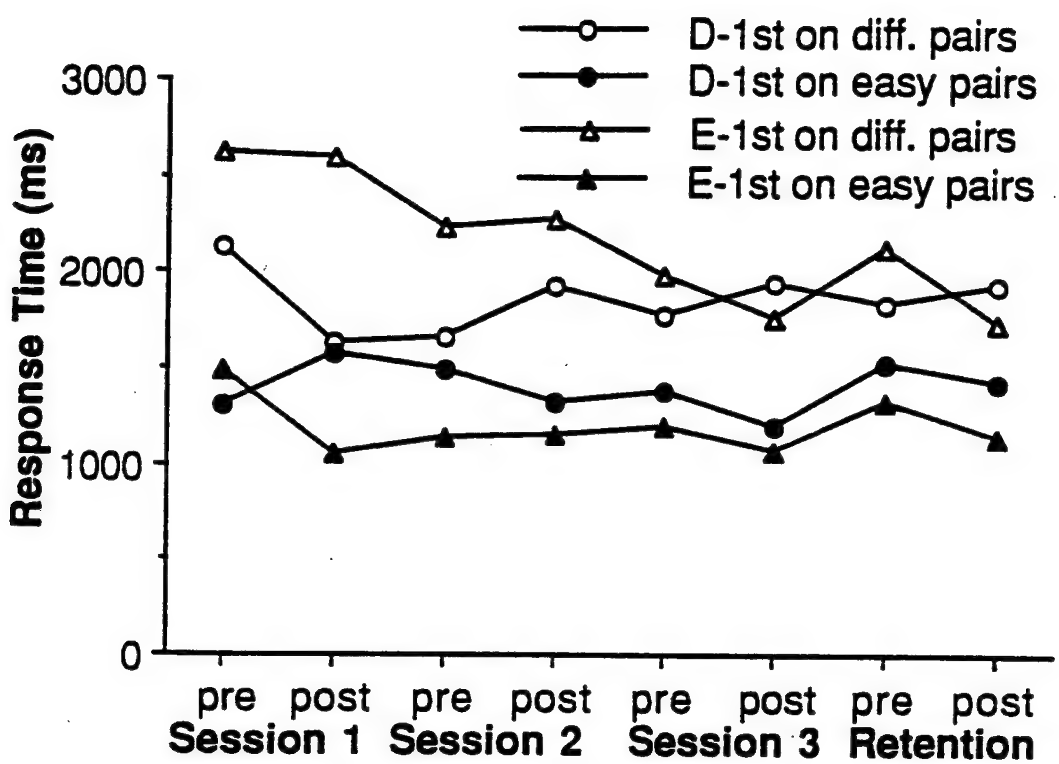

D-1st on diff. pairs
D-1st on easy pairs
E-1st on diff. pairs
E-1st on easy pairs
3000
2000
1000
0
Response Time (ms)
pre
post
pre
post
pre
post
pre
post
Session 1
Session 2
Session 3
Retention

Figure 4

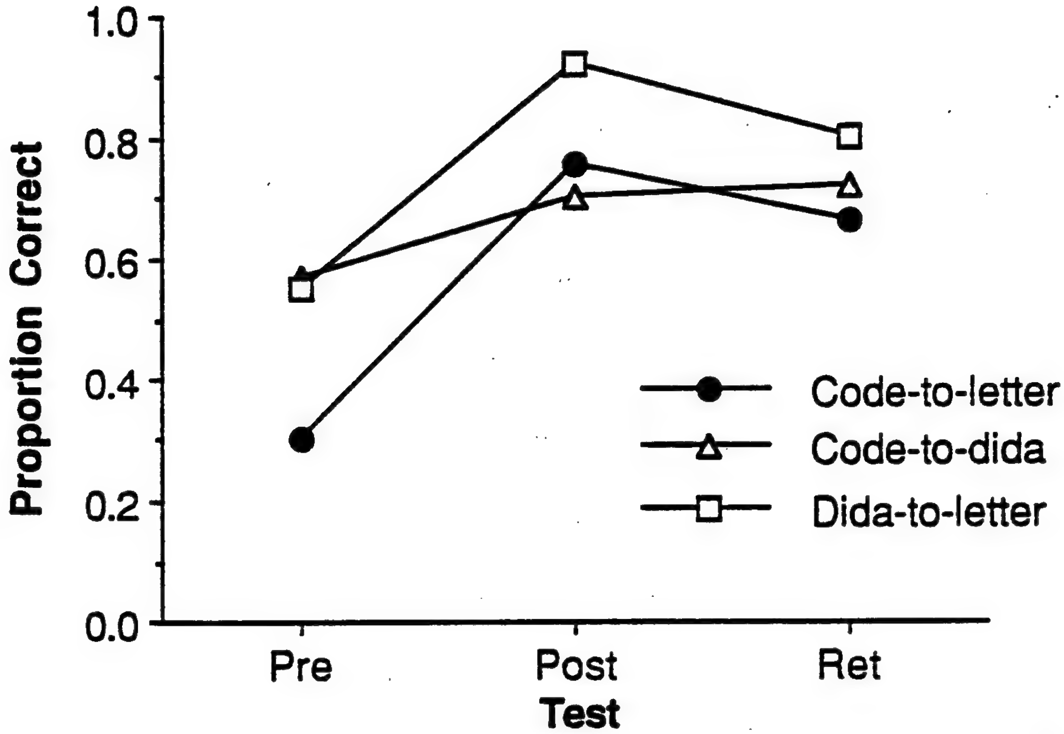

1.0
0.8
0.6
0.4
0.2
0.0
Proportion Correct
Code-to-letter
Code-to-dida
Dida-to-letter
Pre
Post
Ret
Test

Figure 5

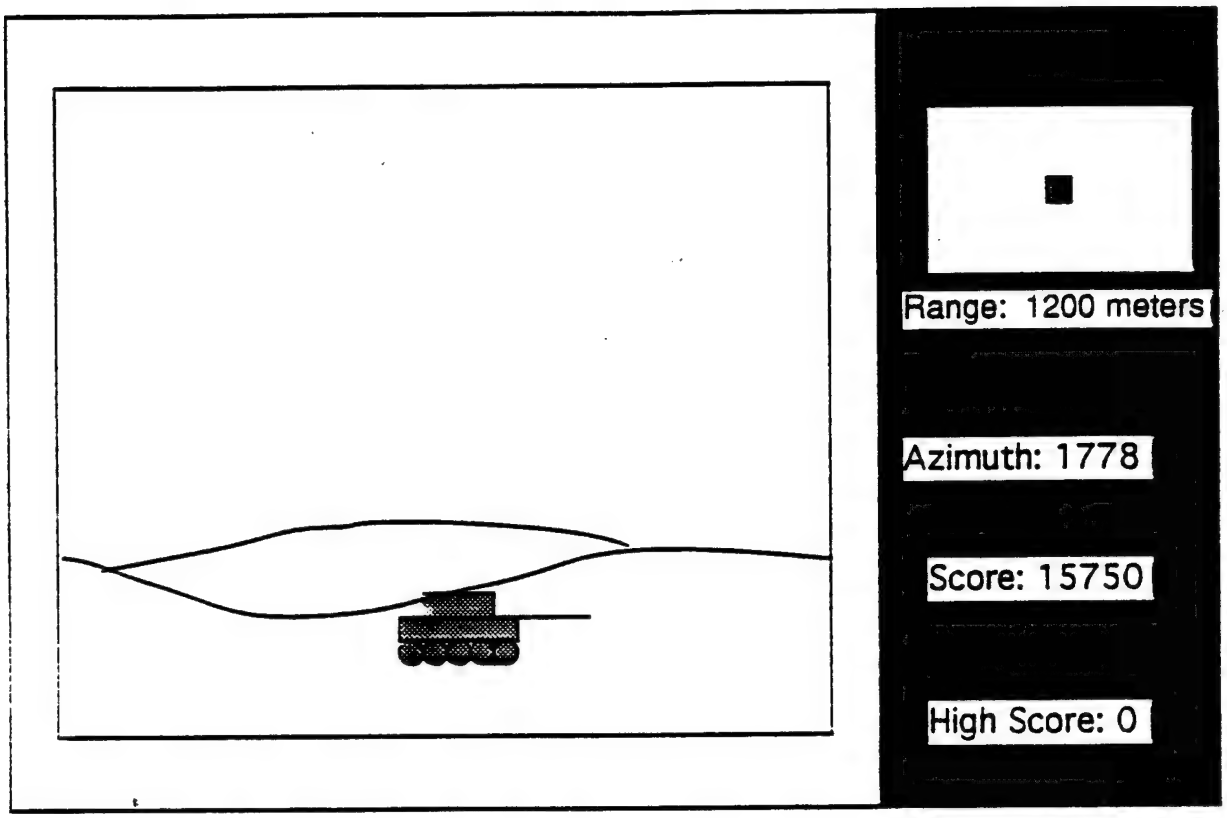


Figure 6

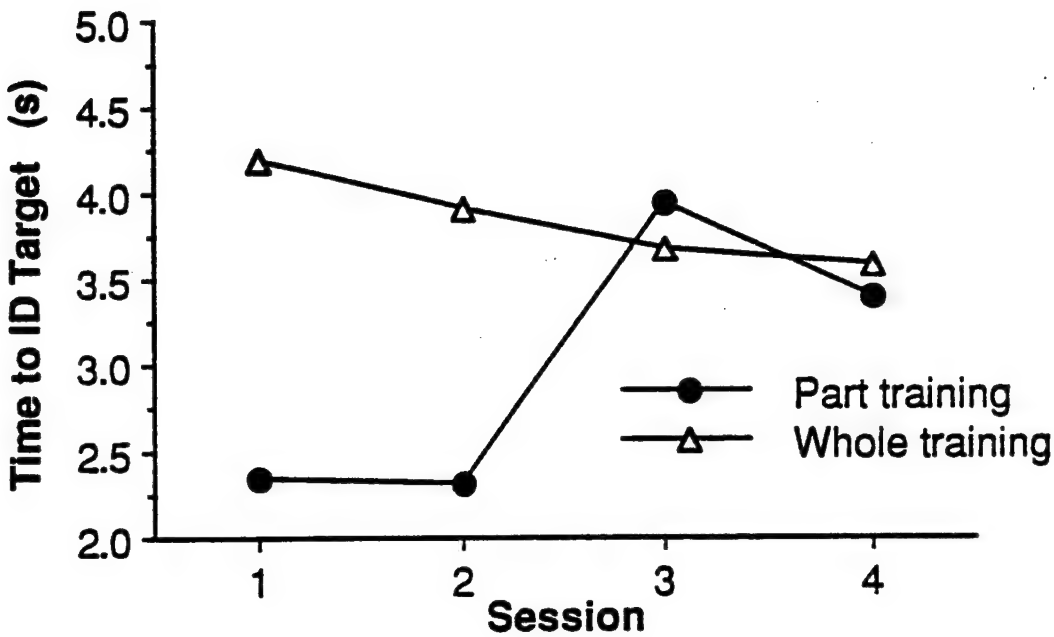

5.0
4.5
4.0
3.5
3.0
2.5
2.0
Time to ID Target (s)
1
2
3
4
Session
Part training
Whole training

Figure 7

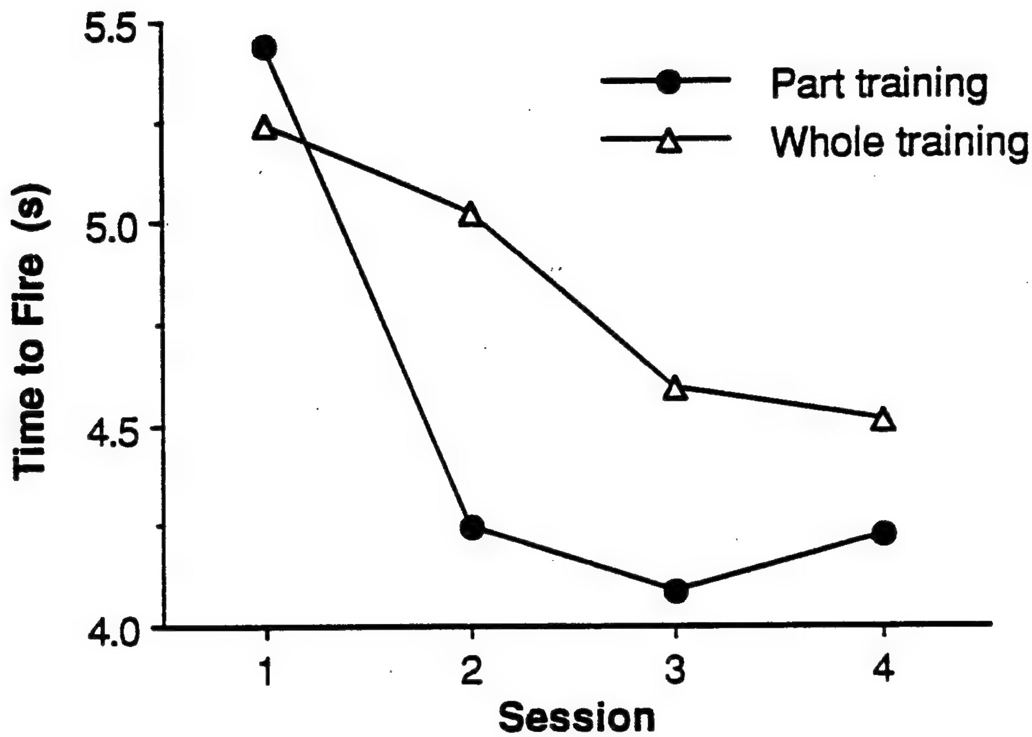

5.5
5.0
4.5
4.0
Time to Fire (s)
Part training
Whole training
1
2
3
4
Session

Figure 8

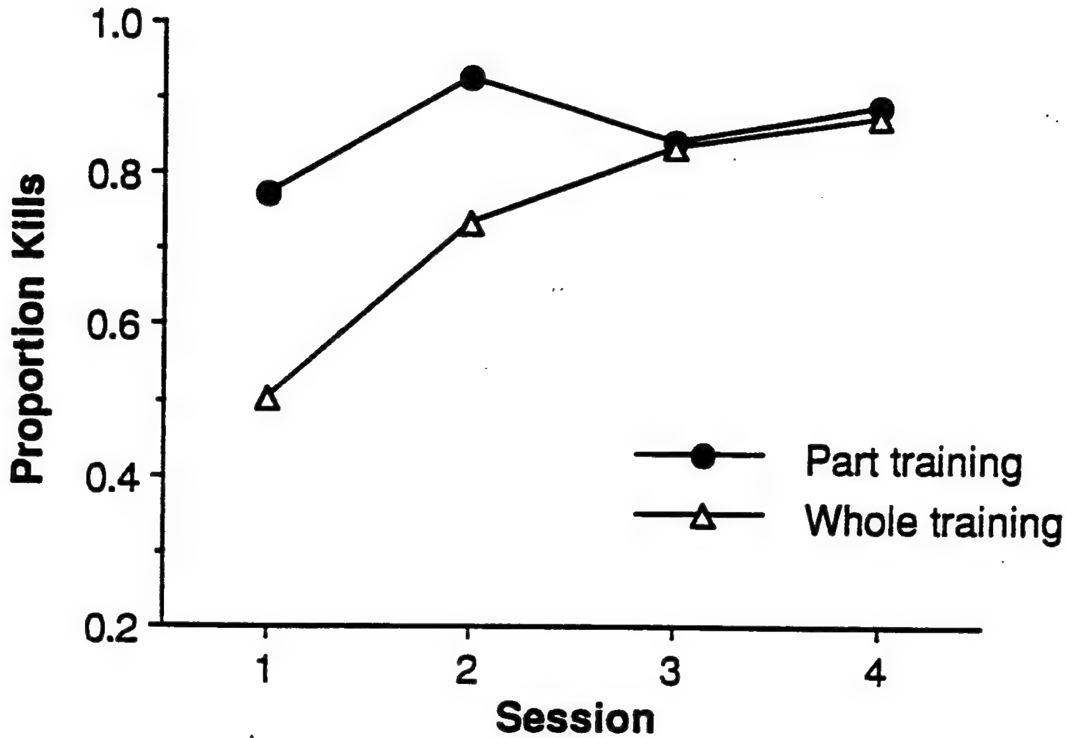


Figure 9

Healy et al., The Long-Term Retention..., The Psychology of Learning and Motivation, Figure #

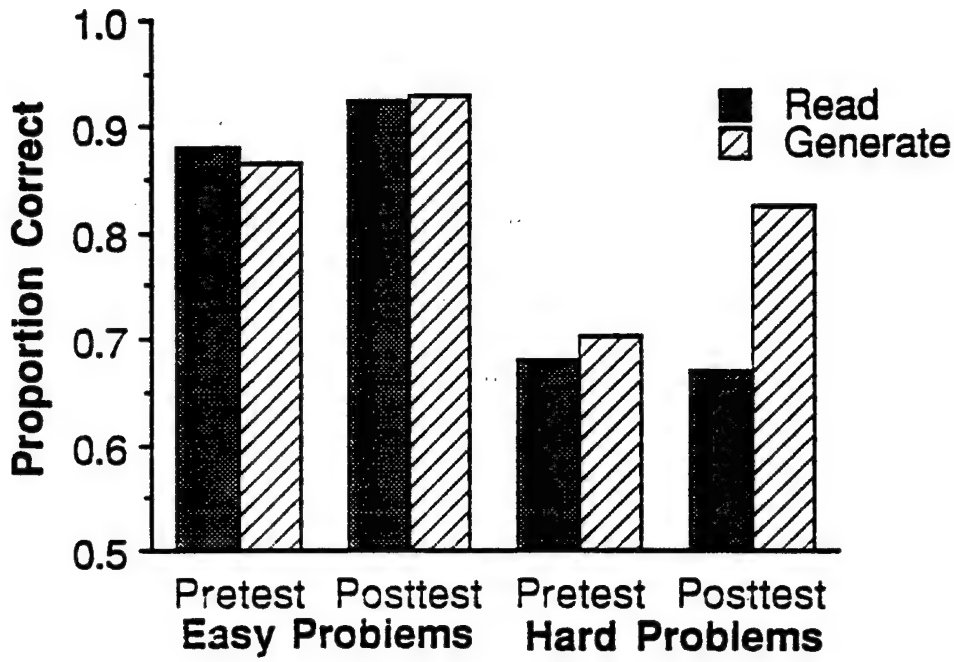

1.0
0.9
0.8
0.7
0.6
0.5
Proportion Correct
Read
Generate
Pretest
Posttest
Pretest
Posttest
Easy Problems
Hard Problems

Figure 10

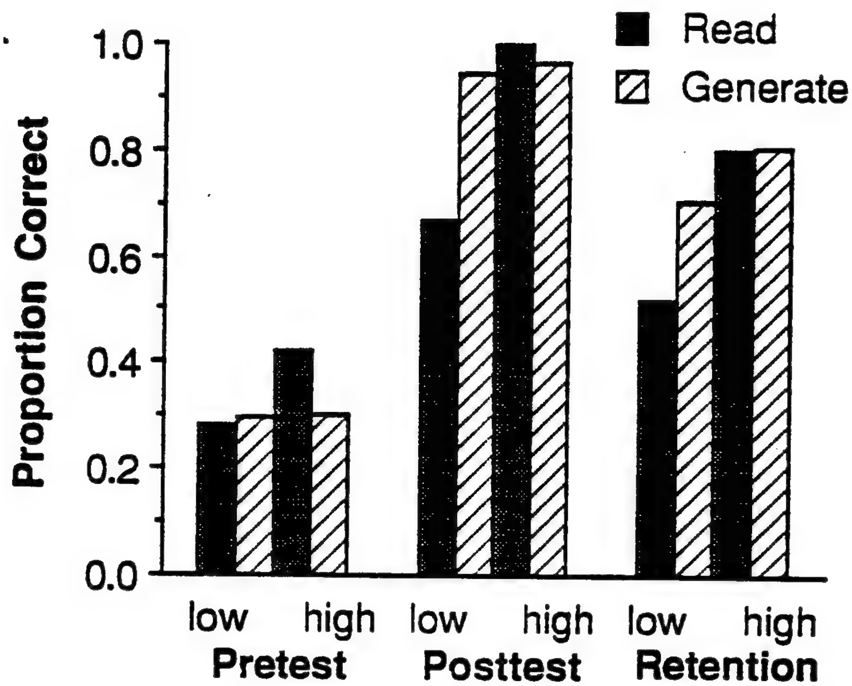

Read
Generate
Proportion Correct
1.0
0.8
0.6
0.4
0.2
0.0
low
high
low
high
low
high
Pretest
Posttest
Retention

Figure 11

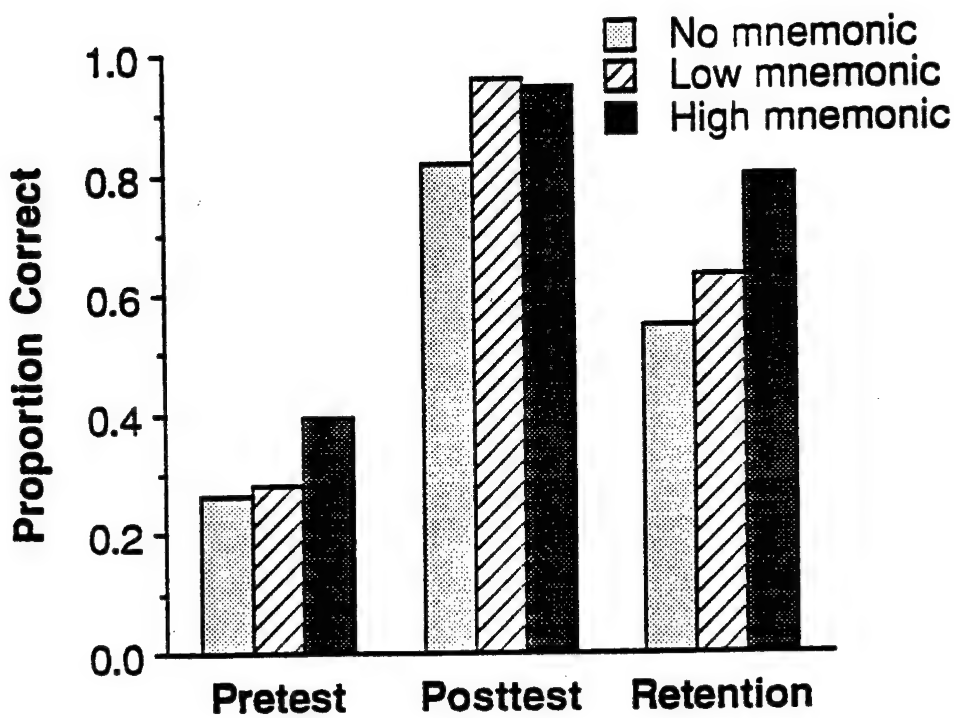

No mnemonic
Low mnemonic
High mnemonic
Proportion Correct
1.0
0.8
0.6
0.4
0.2
0.0
Pretest
Posttest
Retention

Figure 12

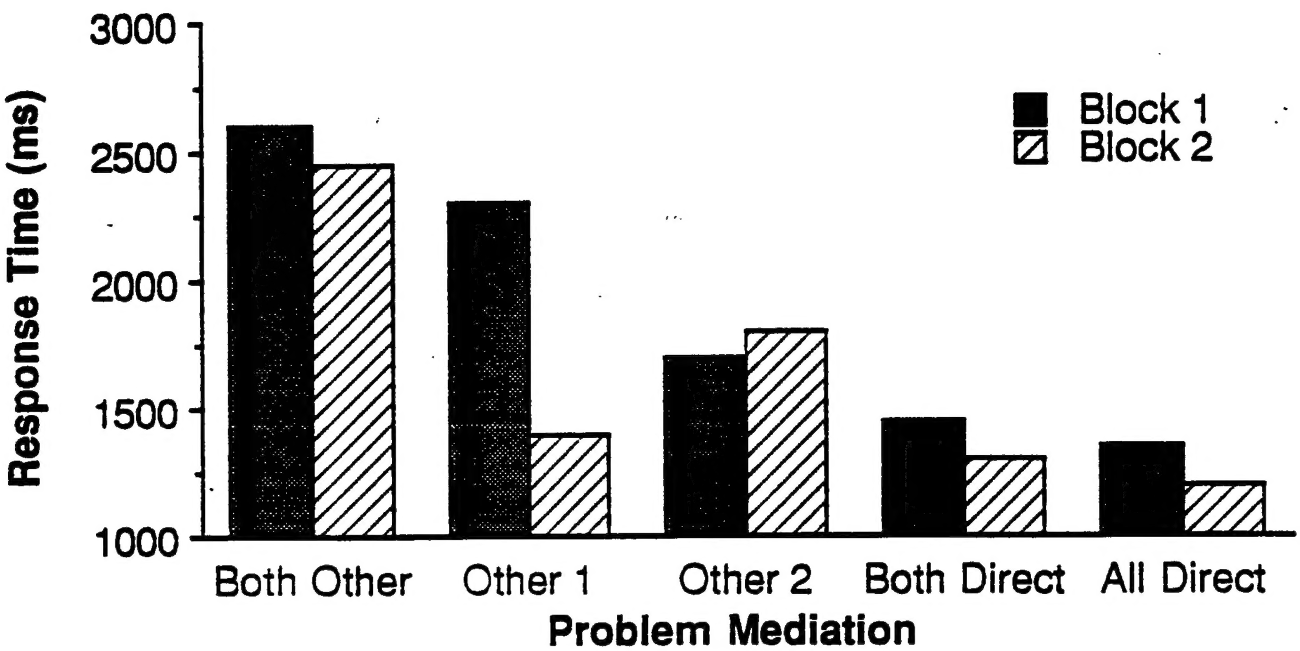


Figure 13

Healy et al., The Long-Term Retention..., The Psychology of Learning and Motivation, Figure #

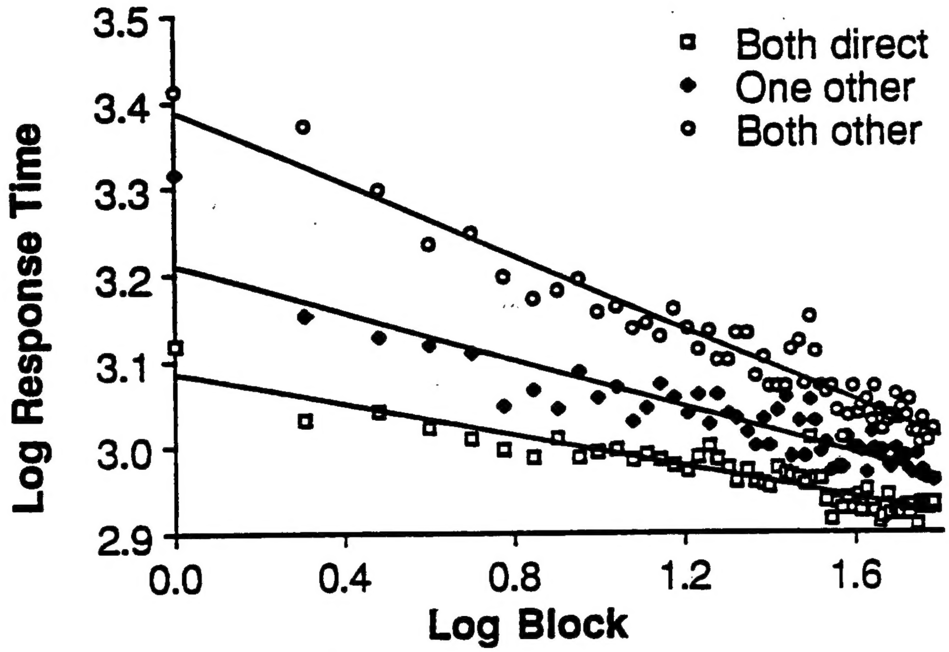


Figure 14

Healy et al., The Long-Term Retention..., The Psychology of Learning and Motivation, Figure

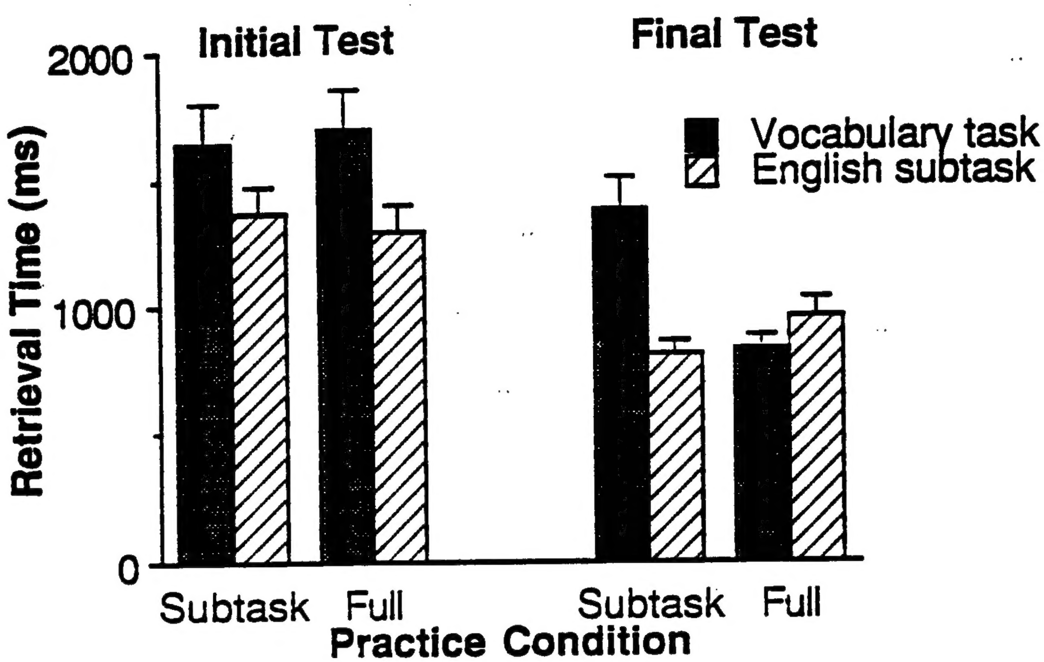

Initial Test
Final Test
2000
1000
0
Retrieval Time (ms)
Vocabulary task
English subtask
Subtask
Full
Subtask
Full
Practice Condition

Figure 15

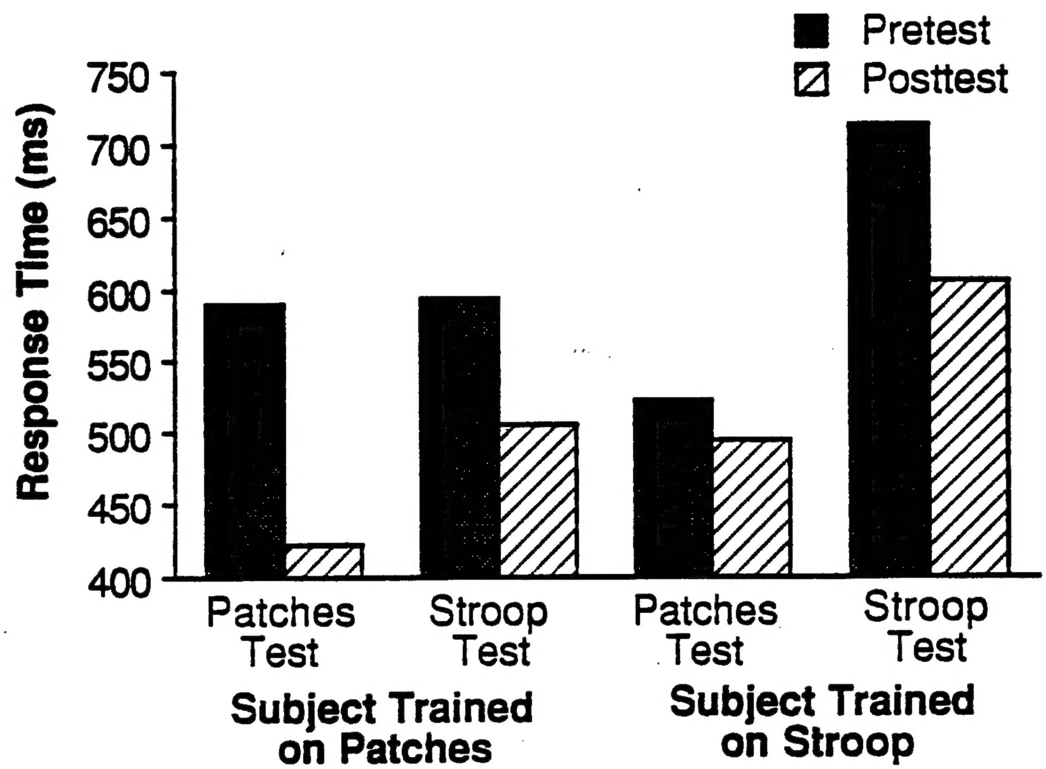

Pretest
Posttest
750
700
650
600
550
500
450
400
Response Time (ms)
Patches Test
Stroop Test
Patches Test
Stroop Test
Subject Trained on Patches
Subject Trained on Stroop

Figure 16